Cell, Vol. 83, 1063–1066, December 29, 1995, Copyright © 1995 by Cell Press

DNA Transposition: From a Black Box to a Color Monitor

Minireview

Nigel D. F. Grindley and Andres E. Leschziner
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut 06520-8114

It is nearly half a century since the classical genetic experiments of Barbara McClintock revealed to her the existence of transposing DNA elements in maize (see McClintock, 1992). The intervening years have witnessed the identification of an ever-increasing number of similar elements from the lowly bacterial insertion sequences to the complex genome of bacteriophage Mu and from the invaluable P element, tool of Drosophila genetics, to the medically devastating HIV retrovirus. Time has also seen dramatic growth in our level of knowledge and understanding of DNA transposition as the geneticists have given way first to the molecular biologists, then to the biochemists, and now to the X-ray crystallographers. In four recent papers, the crystal structures of the catalytic domains of three transposase proteins have been reported, starting in December of last year with the HIV integrase and now with the Mu transposase and the avian sarcoma virus (ASV) integrase (Dyda et al., 1994; Rice and Mizuuchi, 1995; Bujacz et al., 1995, 1996).

The Process of Transposition

Biochemical studies have shown that the enzymes of transposition, the transposases (including retroviral integrases), are responsible for the breaking and joining of DNA during the movement of transposable elements (Mizuuchi, 1992; Katz and Skalka, 1994). In some, perhaps all, cases they are also responsible for bringing together the two element ends in a synaptic complex prior to catalysis of DNA cleavage. During transposition, a transposase performs two distinct and sequential reactions on its DNA substrate (Figure 1). First, following recognition and binding to the transposable element, the transposase acts as a site-specific endonuclease cleaving the DNA to expose the element's 3′-OH ends. A subset of transposases (e.g., Tn7, Tn10) cleave both DNA strands, severing all connections between transposon and flanking DNA; however, only the 3′-OH ends participate in the second phosphoryl transfer (see Mizuuchi, 1992). A duplex target DNA is then brought into the complex and, by direct attack of the 3′-OH ends of the element on the relevant phosphates, is cleaved and joined to the element's ends in a single step. Since the phosphates attacked by the transposon ends are staggered (by 5 bp in the case of Mu and HIV and by 6 bp for ASV), the product contains a short single strand gap at each transposon–target junction; this is subsequently filled in, presumably by host repair functions. Strong evidence that the phosphoryl transfers involved in 3′ processing and strand transfer both occur by a single-step mechanism has been obtained for HIV integrase and Mu transposase using chiral thiophosphates at the cleaved position; each single reaction analyzed resulted in an inversion of the stereochemical configuration of the phosphorothioate (Mizuuchi, 1992). Divalent cations are required for both catalytic steps. However, although Mg^{2+} is expected to be the relevant cofactor in vivo, some of the transposases so far studied only show detectable activity in vitro in the presence of Mn^{2+}.

The chief similarities and differences between the reactions catalyzed by Mu transposase and the retroviral integrases are illustrated in Figure 1. The natural substrate for integrase, IN, is a linear duplex DNA copy of the viral RNA genome with just 2 bp beyond the ends of the transposing unit. In Mu transposition, by constrast, the transposon is embedded in long genomic sequences derived from its bacterial host and is supercoiled. In vivo, strand transfer

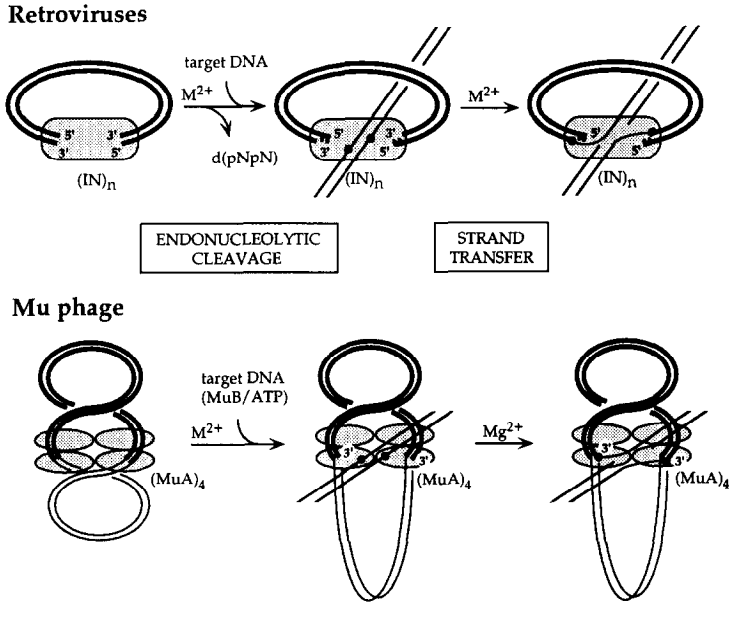

Figure 1. Reactions Catalyzed by the Retroviral Integrases and Mu Transposase

must be a coupled, two-ended reaction (since both ends of an element insert into the same target site), indicating that it occurs within a synaptic complex. For Mu, assembly of this synaptic complex, consisting of a tetramer of transposase, MuA, at the paired ends of Mu, is a prerequisite not only of strand transfer, but also of end cleavage in vitro (Mizuuchi et al., 1992). Synapsis of two ends appears to be important also for 3′-end processing by retroviral integrases (Murphy and Goff, 1992; Kokolj and Skalka, 1995), even though both cleavage and strand transfer steps occur in vitro as uncoupled reactions with single-ended substrates, producing single-end insertion products.

A Transposase Superfamily

The Mu transposase and the retroviral integrases are, at first sight, very different enzymes. The Mu transposase is a 663 residue (75 kDa) protein divisible into several functional domains, the largest of which (327 residues) contains the active site (although recent reports suggest contributions from another domain [Wu and Chaconas, 1995; Yang et al., 1995]) and binds nonspecifically to DNA (see Mizuuchi, 1992). Other domains are responsible for binding specifically to the ends of the transposon, for binding to an enhancer-like sequence, and for interaction with the MuB protein (an activator of transposition). The HIV and ASV integrases are less than half the size of MuA (about 290 residues) and contain three domains: an N-terminal domain of uncertain function but with a His_2–Cys_2 zinc-binding motif, a central catalytic domain (about 160 residues), and a C-terminal domain with DNA-binding activity (see Katz and Skalka, 1994).

Despite their many obvious differences, which probably reflect the differences in lifestyle of each virus, MuA and the retroviral integrases may be related. Extensive and careful sequence comparisons over the last few years have indicated that the integrases are related to many transposases; sequence similarities are confined chiefly to the catalytic domain, and, although quite extensive in some cases (e.g., between the integrases and the IS3 family of transposases), in others they are barely detectable (Fayet et al., 1990; Kulkosky et al., 1992; Doak et al., 1994). Interestingly, Mu transposase was not included in this superfamily of proteins until very recently, and even then the preferred sequence alignments (Baker and Luo, 1994; Kim et al., 1995) have been shown by structural comparisons to be only partially correct (see below). The most notable feature of these conserved segments is a triad of "invariant" carboxylate residues, two aspartates and a glutamate known as the DDE motif. These three conserved residues are well separated in the primary sequence with spacers of about 50–70 residues between the two aspartates and 35 residues (often, but not always) between the second aspartate and the glutamate. Even conservative substitutions at these positions have a drastic effect on cleavage and strand transfer (Kulkosky et al., 1992; Baker and Luo, 1994; Kim et al., 1995). The presence of these conserved carboxylates, their essential nature and the requirement for divalent metal ion during catalysis led to the suggestion that these enzymes may promote phosphoryl transfers by a process similar to the two-metal ion mechanism proposed for the 3′–5′ exonuclease of DNA polymerase I (see Joyce and Steitz, 1994).

Three Transposase Crystal Structures

The X-ray crystal structures of the catalytic core domain of three members of the transposase/integrase family (Figure 2) have both confirmed some predictions and provided some surprises. As has so often been the case in structural biology, additional information came to light with each structure, a strong argument for independent efforts on related problems.

HIV Integrase

The portion crystallized by Dyda et al. (1994) was the catalytic core domain, residues 50–212. Although this domain fails to catalyze end-processing or integration, it does perform a phosphoryl transfer, releasing the viral end from the branched strand transfer product (a reaction called disintegration); thus, the core domain contains an intact catalytic site. The insolubility of the HIV-1 integrase (indeed, of transposases in general) and of the integrase core domain alone appeared to be an almost insurmountable barrier to crystallographic studies. In an effort, clearly successful, to overcome this problem, Craigie and colleagues performed a mutagenic tour-de-force by systemically replacing all hydrophobic residues in the core domain of the integrase. Much improved solubility was obtained with the single substitution of lysine for Phe-185. This mutant retained disintegration activity. Crystals of this mutant domain were grown in the absence of either DNA substrate or divalent cations.

The structure alone, a mixed $\alpha\beta$ structure with a central β sheet, was perhaps a little disappointing, since only the two aspartate residues (Asp-64 and Asp-116) of the DDE triad were visible. These two were close to one another, however, and the third carboxylate, Glu-152, located on a 13 residue disordered region, seemed likely to be located in the same general area, compatible with the proposed role of the three carboxylates in metal binding and catalysis. A remarkable feature of the structure was its striking and unexpected similarity to three other proteins involved in phosphoryl transfers: RNase H (both of Escherichia coli and the RNase H domain of HIV reverse transcriptase) and RuvC, the enzyme that resolves Holliday junctions (Katayanagi et al., 1990; Yang et al., 1990; Davies et al., 1991; Ariyoshi et al., 1994). Moreover, the two invariant aspartate residues of integrase superimposed on two catalytically essential aspartates of RNase H that participate in metal ion binding (Davies et al., 1991; for a discussion of the structural similarities, see Yang and Steitz, 1995a). The obvious conclusion is that all these enzymes use a common catalytic mechanism.

One additional interesting feature of the HIV integrase structure was a substantial interface between monomers related by a crystallographic twofold axis. Dyda et al. (1994) suggest that the large distance between the two active sites (about 35 Å) makes this dimer unsuitable for the concerted strand transfer reaction at the two ends (which occur just 5 bp apart [~ 17 Å in B-DNA] at the target

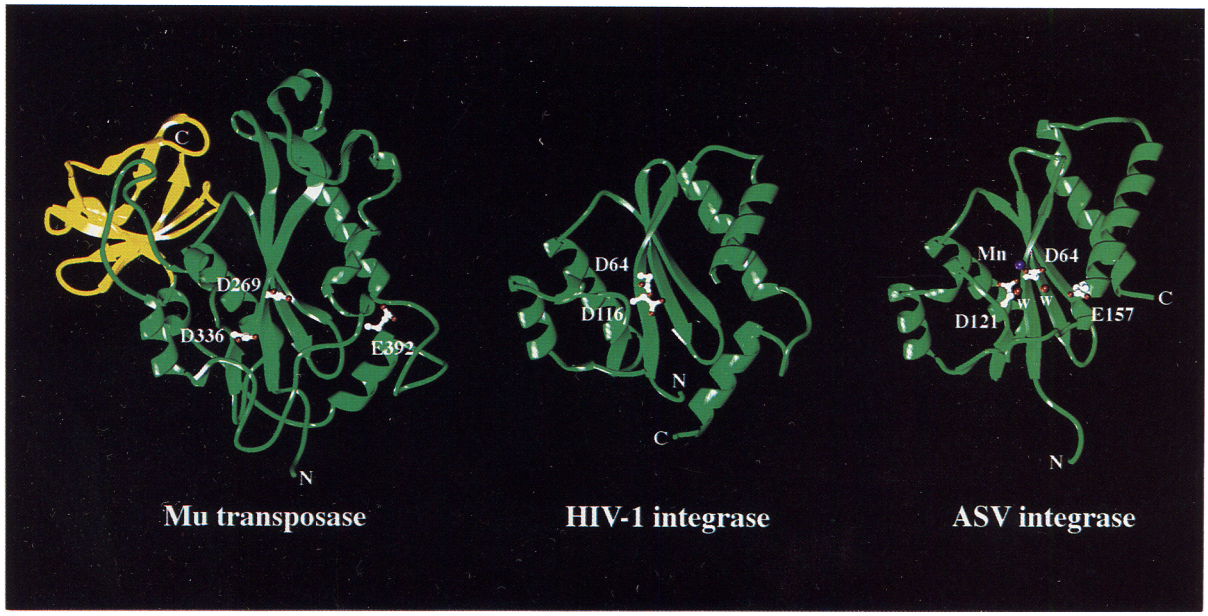

Figure 2. Structures of the Catalytic Domains of Mu Transposase, HIV-1 Integrase, and ASV Integrase

site) but would be compatible with a model involving a tetramer of integrase, with one dimer bound at each end. We note, however, that a comparable situation exists for another recombinase: the dimer of γδ resolvase, which is demonstrably responsible for the coupled cleavages (separated by only 2 bp or about 13 Å) at a single crossover site (Boocock et al., 1995), has widely separated active sites (>30 Å) even in the cocrystal structure with its DNA substrate that was solved recently by Yang and Steitz (1995b).

Mu Transposase

As with the integrases, the crystallized portion determined by Rice and Mizuuchi (1995) was the core catalytic domain. However, in the case of Mu transposase, this domain, residues 248–574, was about twice the size of its integrase counterpart. Several features distinguished the Mu structure. First, it consisted of two subdomains—a 233 residue N-terminal αβ structure with striking structural homology, yet weak sequence similarity, to the integrase domain, and a C-terminal 70 residue β barrel. Electrostatic calculations indicated a large region of positive potential on one surface of the β barrel, and Rice and Mizuuchi argue that it is responsible for the nonspecific DNA-binding activity of the entire core domain.

The second feature of the Mu transposase structure was that all three carboxylates of the conserved DDE triad were visible. Perhaps because divalent metal ions were omitted or because the isolated domain shows no catalytic activity (indeed, the intact transposase is only active in its DNA-dependent tetrameric configuration), the carboxyl groups were not suitably positioned for binding a metal ion cooperatively. The conserved glutamate (invisible in HIV integrase) was on a loop with its side chain pointing out into solvent. Noting that in other homologous structures (HIV integrase and RNase H) this loop is often disordered, Rice and Mizuuchi propose that a simple rearrangement could bring the glutamate into the aspartate-containing active site region; a picture of such a conformational adjustment is provided, perhaps, by the ASV integrase structure discussed below. An important byproduct of the combined structures of Mu transposase and HIV integrase is that they allow a revised, structure-based alignment of the primary amino acid sequences of the two proteins. Just 22 of the 220 Mu residues (160 HIV residues) over the structurally homologous αβ domain are identical between the two proteins (and this number drops by half in a three-way comparison with ASV). The alignment confirms the identities of Asp-269 and Glu-392 and positively identifies Asp-336 as the three members of the Mu DDE motif.

In crystals of the Mu transposase, two regions of the surface appeared to play a role in forming interactions between monomers. However, the details of the interactions varied from one crystal form to another, putting their biological relevance into question. Interestingly, a C-terminal α helix of the catalytic subdomain prevented formation of the dimer interface observed with HIV integrase.

ASV Integrase

Not surprisingly, considering their sequence similarity, the structure of the catalytic domain of ASV integrase (residues 52–207) was very similar to that of the HIV enzyme (although the residues forming the C-terminal α helix of the HIV structure were missing from the ASV domain) (Bujacz et al., 1995). Moreover, the ASV domain crystallized in the same dimeric configuration. In addition to its high resolution, the ASV integrase structure provided an example of a metal ion bound at the active site (Bujacz et al., 1996). Three structures were solved: a metal-free one and two containing either Mg^{2+} or Mn^{2+}. Even in the absence of metal ions, the segment containing the invariant glutamate (disordered in HIV integrase and forming a loop inappropri-

ately positioned for catalysis in Mu transposase) was well ordered, forming the beginning portion of a long α helix, and the glutamate side chain was directed into the active site. There was an extensive network of hydrogen bonds that interconnected the three carboxyl groups via several ordered water molecules. One water interacted with the carboxyl groups of Asp-64 and Glu-157; another was coordinated by Asp-64 and Asp-121. When crystals were soaked in magnesium or manganese, a single metal ion could be seen, replacing the water coordinated by the two aspartates. Metal ion binding was accompanied by only modest changes in carboxylate positions, by far the largest (>1 Å) involving the unliganded glutamate. The presence of only one metal ion could be interpreted as supporting catalysis by a single metal ion, with the third unliganded carboxylate acting as a general base to activate the attacking water (or 3′-OH). However, an open question at this point is whether a second metal-binding site is formed in the presence of the additional negatively charged phosphates that constitute the substrate. The presence of two metal ions in the active site of HIV RNase H with the appropriate 4 Å separation (Davies et al., 1991) inclines us to the view that this class of enzymes will use the two-metal ion mechanism; no doubt time will tell.

These structures of transposase catalytic domains are an important and exciting step toward obtaining a more complete understanding of transposase function. However, many years ago, Jeremy Knowles (see Gutfreund and Knowles, 1967), while acknowledging that a protein structure is essential for understanding its enzymatic mechanism, observed that a model of a horse does not necessarily tell us how fast it can run. He might have added that a picture of just part of a horse with no view of the ground might leave some ambiguity in how the horse even stands up. The structures obtained are a tribute to the persistence and effort of the scientists involved. In the kinetic pathway of understanding mechanism, they put us over the first energy barrier; maybe the rate-limiting step has even been accomplished. Clearly, the next major milestone of a structural nature will be a picture of an intact transposase with metal ion(s) and bound DNA substrate.

Selected Reading

Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994). Cell *78*, 1063–1072.

Baker, T.A., and Luo, L. (1994). Proc. Natl. Acad. Sci. USA *91*, 6654–6658.

Boocock, M.R., Zhu, X., and Grindley, N.D.F. (1995). EMBO J. *14*, 5129–5140.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A., and Skalka, A.M. (1995). J. Mol. Biol., in press.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A., and Skalka, A.M. (1996). Structure, in press.

Davies, J.F., II, Hostomska, Z., Hostomsky, Z., Jordan, S.R., and Matthews, D.A. (1991). Science *252*, 88–95.

Doak, T.G., Doerder, F.P., Jahn, C.L, and Herrick, G. (1994). Proc. Natl. Acad. Sci. USA *91*, 942–946.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., and Davies, D.R. (1994). Science *266*, 1981–1986.

Fayet, O., Ramond, P., Polard, P., Prére, M.F., and Chandler, M. (1990). Mol. Microbiol. *4*, 1771–1771.

Gutfreund, H., and Knowles, J.R. (1967). Essays Biochem. *3*, 25–72.

Joyce, C.M., and Steitz, T.A. (1994). Annu. Rev. Biochem. *63*, 777–822.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1990). Nature *347*, 306–309.

Katz, R.A., and Skalka, A.M. (1994). Annu. Rev. Biochem. *63*, 133–173.

Kim, K., Namgoong, S.-Y., Jayaram, M., and Harshey, R.M. (1995). J. Biol. Chem. *270*, 1–8.

Kukolj, G., and Skalka, A.M. (1995). Genes Dev. *9*, 2556–2567.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P.G., and Skalka, A.M. (1992). Mol. Cell. Biol. *12*, 2331–2338.

McClintock, B. (1992). Nobel lecture. In The Dynamic Genome, N. Federoff and D. Botstein, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 361–380.

Mizuuchi, K. (1992). Annu. Rev. Biochem. *61*, 1011–1051.

Mizuuchi, M., Baker, T.A., and Mizuuchi, K. (1992). Cell *70*, 303–311.

Murphy, J.E., and Goff, S.P. (1992). J. Virol. *66*, 5092–5095.

Rice, R., and Mizuuchi, K. (1995). Cell *82*, 209–220.

Wu, Z., and Chaconas, G. (1995). EMBO J. *14*, 3835–3843.

Yang, J.-Y., Kim, K., Jayaram, M., and Harshey, R.M. (1995). EMBO J. *14*, 2374–2384.

Yang, W., and Steitz, T.A. (1995a). Structure *3*, 131–134.

Yang, W., and Steitz, T.A. (1995b). Cell, *82*, 193–207.

Yang, W., Hendrickson, W.A., Crouch, R.J., and Satow, Y. (1990). Science *249*, 1398–1405.